

AD-A135 565 THE AUTOMATED INFORMATION RETRIEVAL SYSTEM IN THE FIELD 1/1
OF SCIENCE AND SC. (U) FOREIGN TECHNOLOGY DIV
WRIGHT-PATTERSON AFB OH 8 KRYGIER 09 NOV 83
UNCLASSIFIED FTD-ID(RS)T-1527-83 F/G 5/2 NL



END

FILED
104
DTIC



(Handwritten signature)

FOREIGN TECHNOLOGY DIVISION



THE AUTOMATED INFORMATION RETRIEVAL SYSTEM IN THE
FIELD OF SCIENCE AND SCIENCE POLICY--AWION

by

B. Krygier



DTIC
ELECTE
DEC 9 1983
S D D

Approved for public release;
distribution unlimited.



83 12 09 025

A135565

DTIC FILE COPY

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A/1	



FTD-ID(RS)T-1527-83

EDITED TRANSLATION

FTD-ID(RS)T-1527-83

9 November 1983

MICROFICHE NR: FTD-83-C-001364

THE AUTOMATED INFORMATION RETRIEVAL SYSTEM IN THE
FIELD OF SCIENCE AND SCIENCE POLICY--AWION

By: B. Krygier

English pages: 23

Source: Zagadnienia Informacji Naukowej,
Nr. 2(29), 1976, pp. 61-79

Country of origin: Poland

Translated by: SCITRAN

F33657-81-D-0263

Requester: RCA

Approved for public release; distribution unlimited.

THIS TRANSLATION IS A RENDITION OF THE ORIGINAL FOREIGN TEXT WITHOUT ANY ANALYTICAL OR EDITORIAL COMMENT. STATEMENTS OR THEORIES ADVOCATED OR IMPLIED ARE THOSE OF THE SOURCE AND DO NOT NECESSARILY REFLECT THE POSITION OR OPINION OF THE FOREIGN TECHNOLOGY DIVISION.

PREPARED BY:

TRANSLATION DIVISION
FOREIGN TECHNOLOGY DIVISION
WP-AFB, OHIO.

FTD-ID(RS)T-1527-83

Date 9 Nov 19 83

GRAPHICS DISCLAIMER

All figures, graphics, tables, equations, etc. merged into this translation were extracted from the best quality copy available.

THE AUTOMATED INFORMATION RETRIEVAL SYSTEM IN THE FIELD OF SCIENCE AND SCIENCE POLICY--AWION

by Barbara Krygier

The characteristics of the Automated Information Retrieval System for Science--the system's subject scope, task and function. The information source for the system and set organization. Science and science policy thesaurus. The general basis of data processing. The conditions for the functioning of the AWION system and the linking of the system with domestic and foreign information systems in the field of social sciences.

THE SUBJECT SCOPE

As is known, even after many years of discussion, the scope of science has not been unanimously defined. We start from the principle that in order to define information source in science, from the practical viewpoint, we must adopt a broad understanding of such terms as set title for all those theoretical and practical interests, which are the subject of science. The system itself must encompass a broad information base, which concerns scientific subjects, leaving the user the possibility of individual source selection and the

freedom to deal with material closely or only marginally connected with his interests.

Moreover, in accordance with the views of I. Malecky, we accept the fact that studies of science policy and science are concepts which only partially coincide. The difference lies both in their research subjects and in their characteristics^{x/}. Therefore, the term--information retrieval system in the field of science and science policy--has been used in the project materials.

The extension of the field of information, which can be grasped in stages, is connected with this understood subject scope. The introduction, above all, of sources which concern science policy, therefore, was anticipated in the first period of the system's realization.

The subject, which enters into the scope of the system, can in a conventional manner be generally understood by aid of a schema, based on the divisions of the MCNTI (International Center for Scientific and Technological Information) list of classification headings. The expanded points 4, 5, 6 7 and 8 point to the questions preferred from the point of view of the system's subject profile, at least in the first stage of its realization.

The issues of science, enumerated below, cannot be treated as classifications, but rather as the general scope of the problem, which concern documents in regard to the system.

^xMalechi, I.: The problems and range of studies in political science, "Problems of Science", 1974, V. 4/40, p 478-499.

x/ Malecki I.: the Task and scope of studies on science policy.
"Zagadn. naukoznawstwa", 1974 4/40 pp. 478-499.

SCIENCE, ORGANIZATION AND THE PLANNING OF SCIENTIFIC RESEARCH

1. The total problem
2. The theory and methodology of science. Sciometrics
3. The classification of science
4. Forecasting in science
 - 4.1 Forecasting methods in science and technology
 - 4.2 The working out of Forecasts of the development of science and technology
5. The organization and planning of scientific research
 - 5.1 Scientific management
 - 5.2 Organization and activity and institutes and scientific organizations
 - 5.3 The planning and economic calculation of scientific research
 - 5.4 The activity of scientific-research institutes of higher education
 - 5.5 The practical results of research
 - 5.6 International scientific cooperation
6. The economics and efficiency of scientific research
 - 6.1 The economics of scientific research
 - 6.2 The efficiency of the activities of scientific institutions

and design and project bureaus

7. Scientific cadres and their training

7.1 Defining the need for scientific-research workers. The numerical standard for scientific cadres

7.2 Preparing scientific cadres/post-baccalaureates, doctors/

7.3 The selection, positioning and testing of scientific cadres

7.4 The ideological-political education of scientific cadres

7.5 The training of scientific cadres abroad

8. The organization of the activities of scientific workers

8.1 The scientific organization of work. The organization of work stations

8.2 The activities of individual categories of scientific workers

8.3 The remuneration of scientific workers. Material incentives

9. Scientific terminology

10. The history of science and technology

THE SYSTEM'S TASKS AND FUNCTIONS

The basic goal of the automation of information services, specialized and determined by the system's selected subject scope--and therefore, with a relatively limited number of documents from the point of view of automation--is not only the acceleration of

information retrieval, but, above all, the possibility of correlated information retrieval in internal information subsets, modified with regard to structure and content--see "The organization of information sets" p. 68.

This is a question of the ability of systems to be composed in the future of data banks which contain information dealing with:

- the determination of existing and future research priorities,
- the allocation of resources necessary for their realization,
- the planning of scientific cadre development,
- the organization of research institutions,
- movements in the social sciences,
- the creating of conditions favorable to research.

System users will have to include:

- research agencies of the Polish Academy of Sciences,
- universities,
- departmental scientific agencies
- individual scientific workers, occupied with scientific problems,
- the chief authorities of the Polish Academy of Sciences and cadres, who are directing scientific research in other departments.

The task of the AWION system is, above all, to supply its users with current and retrospective bibliographic, documentary, legislative and graphic information, as well as providing copies of source documents.

Below is cited the type of institutional information furnished by the system, and then, the type of information, which answers to the

users' questions, is given in the form of currently accessible data.

THE TYPES OF INFORMATION--OUTPUT

1. Information output

- reviews
- signal information
- synthetic elaboration

2. Bibliographic specifications

- periodic, according to the profile
- retrospective

3. Documentational specifications

- periodic, according to the profile
- retrospective

4. Documentational specifications in the area of legislative information--supplemented by fragments of the texts of legal acts.

5. Numerical data concerning employment in scientific agencies, the number of doctors, the amount of expenditures on scientific research, etc.

6. Data concerning individual personnel/authors/ such as:

- bibliographical works,
- projects, research programs in which given individuals have or are currently participating,
- the titles of doctoral and habilitation dissertations,

--information about trips abroad.

7. Data concerning the structure of scientific agencies, for ex.:

--information about the administration in power in a given agency,

--structural changes,

--the state of employment, etc.

FORMS OF REPLY

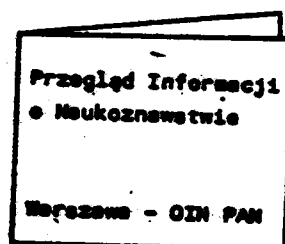
1. Review of information about science, Warsaw--Center of Information Science, Polish Academy of Sciences

Information output:

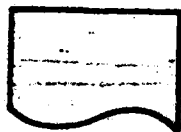
--Reviews

--Signal information

--Synthetic elaboration



2. Machine printers



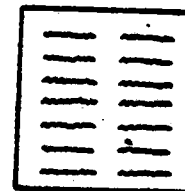
--Bibliographical and documental specifications for continuous inquiries

--Periodic specifications within a subscription framework

3. The photocopying of documents

--Ordered on the basis of PION or specifications

--Ordered within the subscription framework

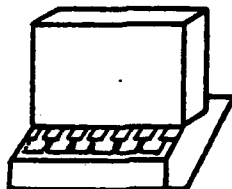


4. Microfiche



--magnetic tape/exchanges, for ex., with collaborating systems

5. The utilization of system services by individual terminals with screen monitors



THE INFORMATION SOURCE

Various types of published/coherent and continuous outputs/ and non-published materials make up the information sources for this system. Among the published sources, we must mention output conveyed by the bookseller and publishing fields, which is very valuable with respect to its rarity and lack of accessibility, disseminated outside the bookseller's field.

Moreover, unconventional documents/for ex. reports, protocols, programs, etc./, which contain the most current data and are the most difficult to obtain, are especially important for the accumulation of complete information.

The utilization of this type of document can be guaranteed only through close cooperation with organizations and agencies which operate in the field of the problems encompassed by the system. And so, for example, the inclusion of documents, which have appeared as a result of many institutions, into the information set of the AWION system is anticipated. These institutions include:

--The Institute for Science Policy and the High School of the Polish

Academy of Sciences and the MNSzWiT,

--The Bureau for the Planning and Coordination of Scientific Research
of the Polish Academy of Sciences,

--The Bureau for Scientific Cadres and Personnel Matters of the Polish
Academy of Sciences,

--The Presidential Bureau of the Polish Academy of Sciences,

--The Science Committee of the Polish Academy of Sciences,

--The Publications Bureau and Library of the Polish Academy of
Sciences.

Given below are the most important types of documents, contained in the
system.

TYPES OF DOCUMENTS CONTAINED IN THE SYSTEM

--Monographs

--Articles from journals

--Bibliographies of bibliographies

--Special bibliographies

--Encyclopedic publications

--Special and terminological dictionaries

--Statistical publications

--Proceedings from congresses and symposia

- Directories
- Indexes
- Promulgative publications
- The programs of congresses and symposia
- Research plans and programs
- Research reports/including unpublished ones/
- Criteria
- Syntheses
- Briefs of legal acts
- Statutes
- Dissertations
- Others

If the size of the document set in the AWION system is in question, it is based on the number calculated for the needs of the SPINES system by the Science Policy Division--UNESCO. It is anticipated to grow by 20 thousand documents a year.

THE ORGANIZATION OF THE INFORMATION SET

The data is introduced into the system in the form of source descriptions with the simultaneous introduction of the source's full text on miniaturized forms/microfiche/.

The source description consists of three parts:

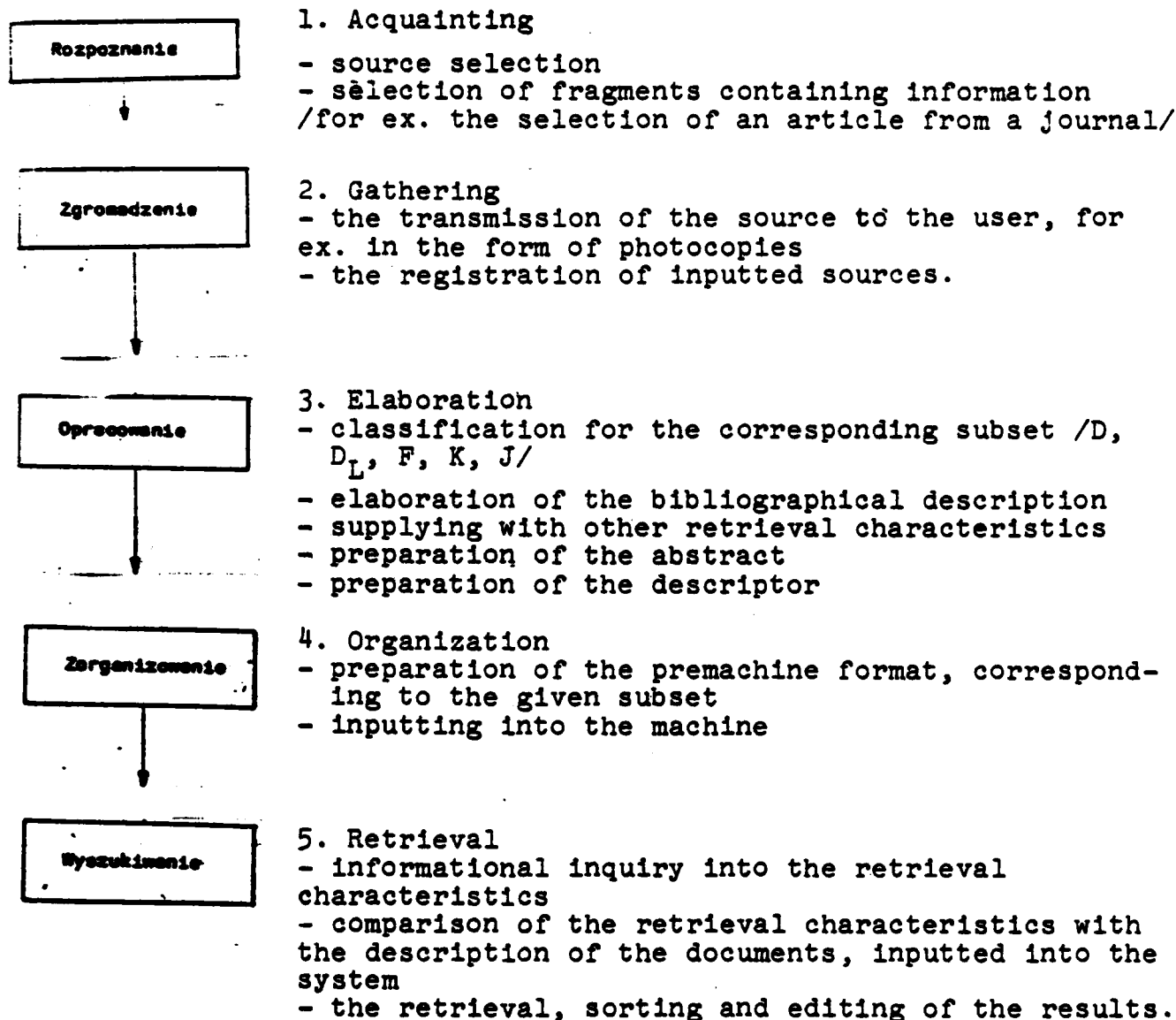
1. Elements of bibliographical description

1. The publication's title in its original language
2. The author's first and last name
3. The publisher and the name of the institution supporting the subject
4. The position number in the instalment of the Przegląd Informacji o Naukoznawstwie [The Information Survey of Science]
5. The abbreviation of the journal's title according to the abbreviation index
6. The abbreviation of the language in which the document is published
7. The abbreviation of the country in which the document is published
8. The year of publication
9. The journal's sequence number/for articles/
10. The abbreviation of the institution in which the document is found
11. The number of pages
12. The designation of the type of document: A--article, B--book, R--report, S--criteria, M--conference materials, P--legal acts, Z--collections of articles, essays, I--directories

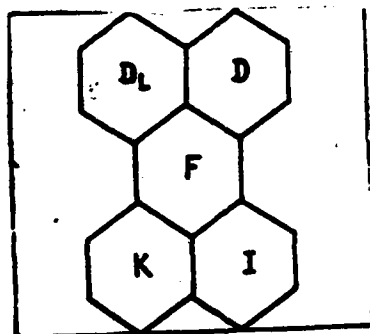
2. Descriptor

3. The description in natural language /summary/

THE OPERATIONS, WHOSE SUBJECT IS THE INFORMATION CONTAINED IN THE SYSTEM



The information contained in the system forms the set which is composed of five subsets, represented in the schema in ill. 1.



Ill. 1 The logical organization of the information set in the AWION system

D--the documentation subset

D_L--the legislative documentation subset

F--the graphic subset

K--the scientific cadres subset

I--the scientific agencies subset

THE SCIENCE AND SCIENCE POLICY THESAURUS^{x/}

The information retrieval language, which makes possible the indexing of documents in the the AWION system and the formulaton of informational inquiries, is the descriptor language and the grammatical systems are presented in the thesaurus.

The following are used for the sources of descriptors for the thesaurus designed in the Center for Science Information of the Polish Academy of Sciences:

--index cards for subject catchwords, used on the basis of the analysis of 10,000 scientific documents listed in the Przegląd Informacji o Naukoznawstwie i Wiadomościach o Nauce [The Information Survey of Science and News about Science],

--the Polish science bibliography,

--the systematization of science project/M. Mazura/,

--the systematization of science terminology, worked out by the Bureau for Scientific Cadres and Personnel Matters of the Polish Academy of Sciences,

--the MCNTI rubricator,

^x A special discussion of thesaureses of science and science polfey, together with the giving of methods for their development will be presented in the next issue of Zagadnien Informacji Naukowej.

--the SPINES thesaurus,

--the OCDE macrothesaurus,

--other sources/for ex., Polish and foreign thesauruses from other fields, indexes for books from the field of science issues, specialist addenda and propositions, etc./.

The present version of the thesaurus contains 3000 descriptors, grouped in the following fields:

A. Chief descriptors

B. The titles of scientific, technical and economic fields

C. The names of titles, scientific degrees, positions and functions for cadres in scientific agencies

D. The names of national scientific-research agencies, committees, associations, etc.

M. Modifiers

G. The names of types of documents

H. Geographical names and the names of languages

K. International organizations

THE GENERAL PRINCIPLES OF DATA PROCESSING

The project of the machine realization of the AWION system was designed on the RIAD 20/OS-1020/, equipped with external tape memory and disc memory with replaceable discs. The system's initiation is planned for the RIAD 32 machine/OS-1030/.

The general schema for the operation of the information retrieval system with the use of a computer is given in ill. 2. This schema also encompasses the concrete variant of such information retrieval systems as the AWION.

A chain structure of the data's recording in the memory has been adopted. Every document description, repeated in the form of a record has the following permanent form: each has bibliographical data, subordinated to retrieval, to each descriptor is ascribed the address of the next record, which contains either the data or the descriptor.

Besides the series of all the records, which correspond to the description of the documents in the machine's memory, is found the series of all the descriptors and values of the bibliographical marks subordinated to retrieval, which contain the following information:

- 1/ the address of the first record, containing the value of the bibliographical mark or the descriptor;
- 2/ the number of records/ that is to say the chain length/, containing the value of the bibliographical mark or the descriptor.

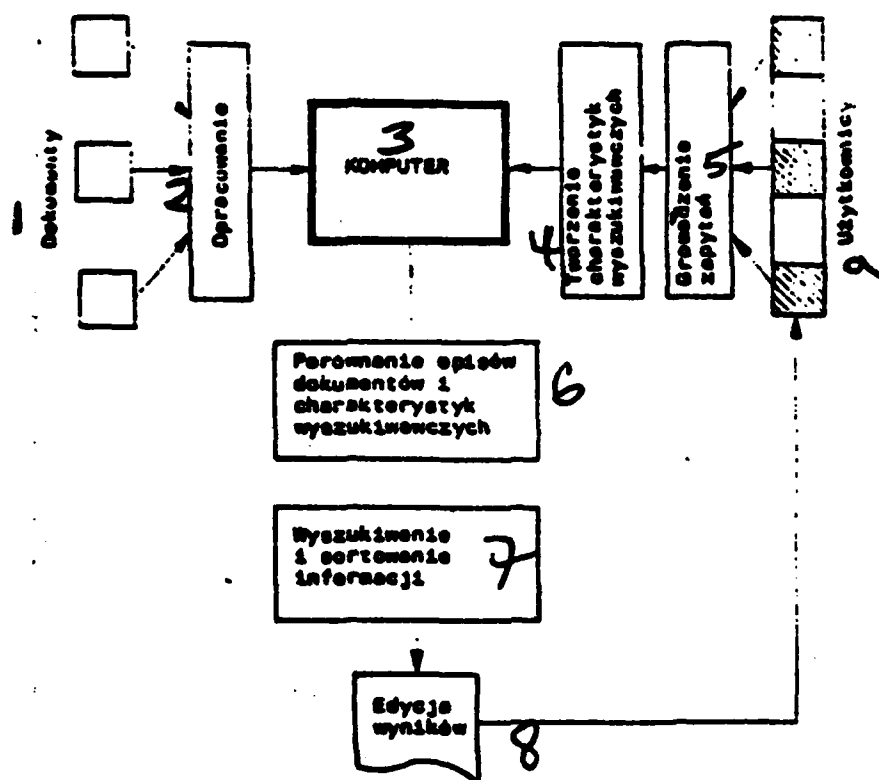
The first information allows "inputting" into the concrete

chain, the second information permits the estimation of the length of such chains from above, without the retrieval operation.

We can differentiate information retrieval, actualization and automatic correction as the basic functions of the system.

In the case of questions, which exist as simple key words, answer retrieval consists of the designation of the address of the initial chain connected to this word and the read out of the records from this chain.

In the case of questions, which exist as the boolean expressions of key words, answer retrieval consists of the read out of the records connected to the shortest chain, which corresponds to a certain descriptor of the question and the subsequent comparison of these records with the given question.



III. 2 the general schema of the operation of the AWION information retrieval system

1. documents, 2. elaboration, 3. computer, 4. formation of retrieval characteristics, 5. gathering of inquiries, 6. comparison of the descriptions of documents and retrieval characteristics, 7. information retrieval and sorting, results edition, 9. users

Two different types of actualization are distinguished in the system:

a/ partial actualization

b/ total actualization

From the point of view of the user, the joint operation in both types of actualization are not differentiated at all. The basic difference concerns the machine organization of the actualized set.

Actualization encompasses the performance of the following operations:

- the connecting of new records to the set of records,
- the elimination of records from the set,
- the connecting of new descriptors/chosen from among those already existant/ to the description,
- the elimination of descriptors from the descriptor,
- the connecting of new descriptors to the set of descriptors,
- the elimination of descriptors from the set of descriptors,
- the elimination of the values of bibliographical marks,
- the modification of the values of bibliographical marks,
- the set's reorganization in the TM,
- the elimination of marks,
- the modification of marks.

Automatic correction is built into the system. This concerns simple errors committed by the system's user, which are caused by the machine's inaccuracies. The descriptors of the dictionary, included in the question's description, are subject to automatic correction.

One of the auxiliary functions of the operating system is the compiling of statistics, allowing the optimization of the system's operation. The compiling of the following statistics are

anticipated:

- 1/ the frequency of the individual types of retrieval errors made in the automatic correction of descriptors,
- 2/ the frequency of the cancellation of specific descriptors,
- 3/ the frequency of the cancellation of specific document descriptions,
- 4/ the frequency of the cancellation of specific summaries,
- 5/ the frequency of the cancellation of specific elements of the bibliographic description during retrieval,
- 6/ the frequency of the errors made in the chosen elements of the bibliographic description, such as: the name of the author, publisher or institution, which sponsors the subject.

The following programs are prepared within the framework of programming in the first sequence:

- the data base loading program
- the publications composition program
- the information retrieval program
- the dictionary retention program

CONDITIONS FOR THE SYSTEM'S FUNCTIONING

The key problem for the realization of the projected system is the securing of an information base, so that the information introduced into the system coincides with at least 80 percent of the

subject scope adopted in the basic assumptions.

If this concerns published documents/journals, books, legal acts, etc./the increase of the input of these sources must be guaranteed by the corresponding organization of purchases, subscriptions and exchanges.

The procurement of unconventional documents/research plans and programs, unpublished reports, syntheses, specifications, projects of legal acts, etc./ will demand the support of agreements on cooperation, the engagement of the authorities, who will accept the responsibility to make available unpublished sources to the system.

Besides the input of information sources, it is necessary to secure the possibility of their rapid and competent creation according to the basic principles and in accord with the instructions for the formation of documents entered into the system/the so-called pre-machine format/.

As experience with many functioning information systems/both domestic and foreign/ has shown, documents can be created well only by specialists from the given field, and then by the system's users. Therefore, it is impossible for a system to function satisfactorily without the activity and cooperation of the user. So, for example, the All-Union Institute for Scientific and Technical Information chooses and creates documents to order for the Institute's informational publications.

An important matter for the proper functioning of a system is solving the problem of protecting information. The designing of

efficient protection mechanisms is included in the work on the technical and organization projects.

x

The AWION system is being designed as the nucleus of a scientific information subsystem of the Polish Academy of Sciences, subordinated to the pan-Polish SINTO /Scientific, Technical and Organizational Information System. The project of the AWION system was designed by the Workshop for Automated Information Processes of the Scientific Information Center of the Polish Academy of Sciences in cooperation with ZETO in Bialymatok and with the collaboration of the Insititute of Logic affiliated with the University of Warsaw in Bialymstok^{x/}.

The problem of linking up with other national subsystems, which encompass information sets about related subjects or have to serve the same circle of users, which are jointly supplemented, is considered in the project principles of the AWION system. This especially concerns the SYNABA system for finished and in-progress research, which was designed by the Institute for Scientific, Technical and Economic Information, and the MAGISTER system, which also encompasses information about scientific-research cadres.

Certain forms of international cooperation are also foreseen within the framework of the information system of the CMEA countries, for ex. with the MISON system/the International Information System for Social Science of the Socialist Countries/, with the system of the

Special mention of the problems connected with the organization and functioning of the system is contained in the project material: in the introductory, technical and use proposals.

Academy of Sciences of the USSR--ASSISTENT, with the system of the Bulgarian Academy of Sciences--NAUKA. Within the framework of the UNISIST program for international cooperation, and especially with the SPINES system/Science and Technology Policies Information Exchange System/cooperation has hitherto only been in the field of the method of system's design, for example, the formation of a thesaurus, indexing principles, etc. When the SPINES system begins to be used close principles of collaboration will be finally harmonized.

LITERATURE

1. Dabrowski M. A mathematical model of a document retrieval system, in Wspolczesne problemy zarzadzania [Modern Problems of Management]. Warsaw: PWN 1974 p. 30-42
2. Darczewski J., Chodorowski M., Michalweicz M. Automated Correction. Zagadnienia Informacji Naukowej 1974 nr. 1/24/ p. 51-79
3. Wojcik J. Information Sources in Science. Waraw 1975. typescript 24 p.
4. Zadrozny S. Some Problems of Information Retrieval. Zagadnienia Informacji Naukowej 1976 nr. 1/28/ p. 73-97

END

FILMED

1-84

DTIC